

Research Article

A Medical Image Classification Model Based on Adversarial Lesion Enhancement

Bing Zhang ¹ and Xu Hu²

¹*School of Information Engineering, Tianjin University of Commerce, Tianjin, China*

²*School of Economics and Management, Tsinghua University, Beijing, China*

Correspondence should be addressed to Bing Zhang; zhangbing@tju.edu.cn

Received 21 April 2021; Accepted 17 May 2021; Published 28 May 2021

Academic Editor: Wenbing Zhao

Copyright © 2021 Bing Zhang and Xu Hu. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

With the development of Artificial Intelligence, the auxiliary diagnosis model based on deep learning can assist doctors to a certain extent. However, the latent information in medical images, such as lesion features, is ignored in most of the traditional methods. The extraction of this information is regarded as a learning task within the network in some recent researches, but it requires a large amount of fine-labeled data, which is undoubtedly expensive. In response to the problem above, this paper proposes an Adversarial Lesion Enhancement Neural Network for Medical Image Classification (ALENN), which is used to locate and enhance the lesion information in medical images only under weakly annotated data so as to improve the accuracy of the auxiliary diagnosis model. This method is a two-stage framework, including a structure-based lesion adversarial inpainting module and a lesion enhancement classification module. The first stage is used to repair the lesion area in the images while the second stage is used to locate the lesion area and use the lesion enhanced data during modeling process. In the end, we verified the effectiveness of our method on the MURA dataset, a musculoskeletal X-ray dataset released by Stanford University. Experimental results show that our method can not only locate the lesion area but also improve the effectiveness of the auxiliary diagnosis model.

1. Introduction

In December 2012, a study [1] showed that musculoskeletal diseases, such as arthritis and back pain, are the second-leading cause of disability as well as the fourth-leading factor to the health of the world population, affecting more than 1.7 billion people worldwide. According to data from the World Health Organization [2], there are more than 150 diseases caused by the musculoskeletal (exercise) system. Although diseases generated by different reasons (for example, exercise, genetics, or poor lifestyle) have little difference in visual perception, there is a big difference among the disease types, so it requires different diagnosis and treatment options. Therefore, the detection of musculoskeletal abnormalities appears particularly important. The abnormalities of musculoskeletal are mainly reflected in the basic diseases of bones, joints and soft tissues [3, 4]. Among them, basic bone diseases include osteoporosis, osteomalacia, bone destruction, osteosclerosis, periosteal proliferation, chondral

calcification, osteonecrosis, bone deformity, etc. Basic joint diseases include swelling of joint, destruction of joint, degeneration of joint, ankylosis of joint, dislocation of joint, etc. Basic soft tissue diseases include soft tissue swelling, soft tissue mass, myophagism, etc.

Bone is the tissue with the highest density in the human structure, which has a clear contrast with surrounding tissues. Meanwhile, there also exists an obvious contrast between the cortex and cancellous bone of the bone itself so that the conventional X-ray examination can be used for general bone diseases diagnosis. In addition, due to the advancement of imaging technology and the upgrading of imaging equipment, various hospitals have produced a large amount of medical imaging data, and these precious data are helpful for many researches. Therefore, it is of great research significance to use computer-aided diagnosis technology to quickly and accurately classify musculoskeletal diseases based on a large number of existing medical images.

Many machine learning methods have been applied to medical image data classification tasks, including K-Means clustering [5, 6], decision tree [7], support vector machine [8], and random forest [9]. However, the number of features extracted by traditional machine algorithms is limited, and only artificially preset features can be classified. Therefore, traditional machine learning algorithms perform poorly on medical image classification. With the continuous development of deep learning technology, deep learning algorithms, including CNN (convolutional neural networks) [10] and GAN (generative adversarial networks) [11], are widely used in classification tasks. Through the fine-grained annotation from a large number of professional doctors, the CNN model can automatically extract features and can extract more feature information through a deeper and wider network architecture. But the difficulty of medical image annotation makes it more expensive to improve CNN model performance. In addition, most of the existing CNN models use benign and malignant labeling information as the last layer of supervision information on the network, and they cannot dig out more hidden features through coarse-grained benign and malignant labeling to reuse and merge key features.

Based on the problems above, this paper proposes an Adversarial Lesion Enhancement Neural Network for Medical Image Classification (ALENN), which automatically recognizes the lesion area in the image only through the supervision of category annotation, and enhances and optimizes the prediction accuracy of the auxiliary diagnosis model. This method is a two-stage model. The first stage is a structure-based lesion adversarial inpainting module, which is used to repair the lesion area in the image; the second stage is a lesion enhancement classification module, which is used to identify the lesion area and apply the data after the enhanced lesion to assist the modeling process of the diagnostic model. The core of the first stage is structural information, which represents the relatively fixed semantics of human body structure in medical images. We believe that better restoration results can be obtained by splitting the image restoration process into structural semantic restoration and texture detail restoration. The core of the second stage is sliding window. Through the sliding of the occluded area, the most significant abnormal area in the image can be found. Finally, we verified the effectiveness of this method on the MURA dataset [12], a musculoskeletal X-ray film dataset released by Stanford University.

The related works of traditional and latest auxiliary diagnosis as well as the related researches on the MURA dataset are introduced in the second section of this paper. The method proposed in this paper is introduced in the third section. The experimental results and analysis are shown in the fourth section. The summary and prospects of the work are given in the fifth section.

2. Related Work

CNN-based medical image analysis method has shown excellent performance in many challenging tasks (disease classification [13], lesion detection [14], fine-grained lesion

segmentation [15]), among which it has deepest research and is most widely applied in disease classification. X-rays examination is one of the most common radiology examinations in the clinical diagnosis of chest diseases. How to combine a large number of existing medical images with the rich clinical experience of professionally trained radiologists is of great significance for the diagnosis of chest diseases. Wang et al. [16] proposed a novel text-image embedding network (TieNet) to extract images and corresponding text representations. It employs an end-to-end trainable CNN-RNN architecture embedded with a multi-level attention model to highlight important image regions and their corresponding text words, and then use image features and text embeddings extracted from related reports to classify the chest X-rays. Coronary angiography is the gold standard for computer-aided diagnosis (CAD), so it is essential to describe in detail the position and the degree of stenosis through coronary angiography for classifying the severity of CAD. Wang et al. [17] used recursive capsule network (RCN) to extract the semantic relationship between clinically named entities in the coronary angiography text, so as to automatically find out the maximum stenosis degree of each lumen, and finally inferred the coronary artery severity according to the improved Gensini method.

The MURA dataset is the largest public musculoskeletal image dataset available currently. Many scholars have conducted numerous experimental studies on this dataset, including the use of traditional machine learning methods and deep learning algorithms. Among them, Pawan et al. [18] used support vector machine (SVM), linear SVM, logistic regression, and decision tree algorithms to detect musculoskeletal anomalies, and introduced a gray-level co-occurrence matrix (GLCM) to preprocess the original musculoskeletal images, comparing them on five evaluation indicators of sensitivity, specificity, precision, accuracy, and F1-score in the end. Pranav et al. [19] designed a 169-layer baseline model DenseNet to solve the musculoskeletal abnormality detection. On the MURA data and above, DenseNet's performance is lower than the worst radiologist in 5 of the 7 basic studies, and the performance of the overall model is also lower than the best radiologist. Subsequently, Dennis et al. [2] proposed the idea of ensemble learning to integrate the well-trained classification models DenseNet201, MobileNet and NASNet-Mobile, and designed the ensemble200 model. Finally, its Cohen Kappa score is 0.66, which is lower than the DenseNet169-layer model. But the F1-score of ensemble200 is better than that of DenseNet model, and the Cohen Kappa score variance of different parts is even lower. Due to the sensitivity of CNN to extract features, when the image undergoes transformations, such as rotation and misalignment, the performance of CNN recognition will be greatly reduced. To solve this problem, SAIF et al. [20] introduced the capsule network architecture. Through its powerful dynamics routing mechanism, an output vector containing many feature information (including spatial direction, vector amplitude) can be obtained to achieve abnormal detection of musculoskeletal X-ray films, and this method only applies to a small number of network layers to use 169 densely connected CNNs Kappa

coefficient of the model. Although the CNN model has shown good performance on many datasets, it performs poorly on some subsets. Luke et al. analyzed that the reason is that the model cannot fully describe the complete changes in the dataset, and he defined this problem as hidden layering. Therefore, Luke et al. [21] evaluated several possible techniques for measuring the effect of hidden stratification and found that hidden stratification can occur in unrecognized low prevalence, low label quality, subtle distinguishing features, and it can lead to more than 20% relative performance difference of the models.

3. Method

Due to the inadequate utilization of lesion information by current deep learning models based on medical image datasets (e.g., MURA) and the inexplicability of deep learning itself, there is still much room for the reliability and credibility of auxiliary diagnosis models to improvement. Aiming at the problems above, this paper proposes an Adversarial Lesion Enhancement Neural Network for Medical Image Classification (ALENN). The method includes two main modules: a structure-based lesion adversarial inpainting module and a classification module based on lesion information fusion. The overall two-stage structure diagram is shown in Figure 1. The LE represents the lesion enhancement.

3.1. Structure-Based Lesion Adversarial Inpainting Module.

The X-ray image of the elbow reflects the density difference among the different tissues of the elbow, and the density distribution in the negative data is obviously different from that in the positive data, as shown in Figure 2. The difference of the distribution is reflected in the lesion area in the positive data. This paper assumes that, in the positive data, the tissue density distribution outside the lesion area is similar to the negative data. Therefore, this distribution difference can be fully utilized to find the location of the lesion in the image, thereby providing additional information for the fusion of the lesion area. If there is a generator that only uses negative data for training, that is, it fits the distribution of negative data under ideal conditions, it can theoretically restore the lesion area in the positive data to the data distribution of normal tissue. In other words, the restored image can be regarded as a piece of pseudo-negative data with a distribution similar to the negative data. Hence, this module is based on the idea of the generative adversarial network to generate positive data into negative data, and regard the relative error between the data as important information for subsequent enhancement of the lesion area.

At the same time, in order to better restore the original semantic information of the image, considering that medical images often contain relatively fixed human structural features-bones and muscles, this module extracts the structural information of the data so that the model pays more attention to the relatively fixed structure, not limited to susceptible texture information. Consequently, inspired by the research [22], this paper adds the models generation and

learning process of structural information on the basis of the generative adversarial network.

3.1.1. Structure Information. For the extraction of structure information, we assume that the image is composed of structure and texture and uses the relative total variation (RTV) in [23] to distinguish the structure and texture in the image. The input image is defined as I , the pixel index in the image is defined as p , and the structure in the image is defined as S . Consequently, the process of using a secondary penalty to strengthen the structural similarity between input and output in the TV-L2-based model can be expressed as

$$\arg \min_s \sum_p \left\{ \frac{1}{\lambda} (S_p - I_p)^2 + |(\nabla S)_p| \right\}, \quad (1)$$

where ∇ represents the first-order difference operation, $(S_p - I_p)^2$ is used to maintain the similarity between the structure and the input image, and $|(\nabla S)_p|$ can be divided into two directions: x direction and y direction:

$$\sum_p |(\nabla S)_p| = \sum_p \left\{ |(\partial_x S)_p| + |(\partial_y S)_p| \right\}. \quad (2)$$

However, the authors of [23] found that the total variation regularizer has limited ability to distinguish between strong structural edges and texture. Therefore, in order not to target a certain type of texture structure image, the overall task can be rewritten as

$$\arg \min_s \sum_p (S_p - I_p)^2 + \lambda \cdot \left(\frac{G_x(p)}{N_x(p) + \varepsilon} + \frac{G_y(p)}{N_y(p) + \varepsilon} \right), \quad (3)$$

where λ is the weight and ε is a small positive number, which should avoid being divided by zero. $G(p)$ represents a general pixel-wise windowed total variation measure and $N(p)$ represents a novel windowed inherent variation, which can be expressed as

$$\begin{aligned} G_x(p) &= \sum_{q \in R(p)} g_{p,q} \cdot |(\partial_x S)_q|, \\ G_y(p) &= \sum_{q \in R(p)} g_{p,q} \cdot |(\partial_y S)_q|, \\ N_x(p) &= \left| \sum_{q \in R(p)} g_{p,q} \cdot (\partial_x S)_q \right|, \\ N_y(p) &= \left| \sum_{q \in R(p)} g_{p,q} \cdot (\partial_y S)_q \right|, \end{aligned} \quad (4)$$

where q is the index of all pixels in a square area centered on point p and $g_{p,q}$ is a weighting function defined according to spatial affinity. However, due to the fact that objective function is non-convex and its solution cannot be obtained directly, a numerically stable approximation of the solution needs to be obtained by decomposing the nonlinear part and the quadratic part.

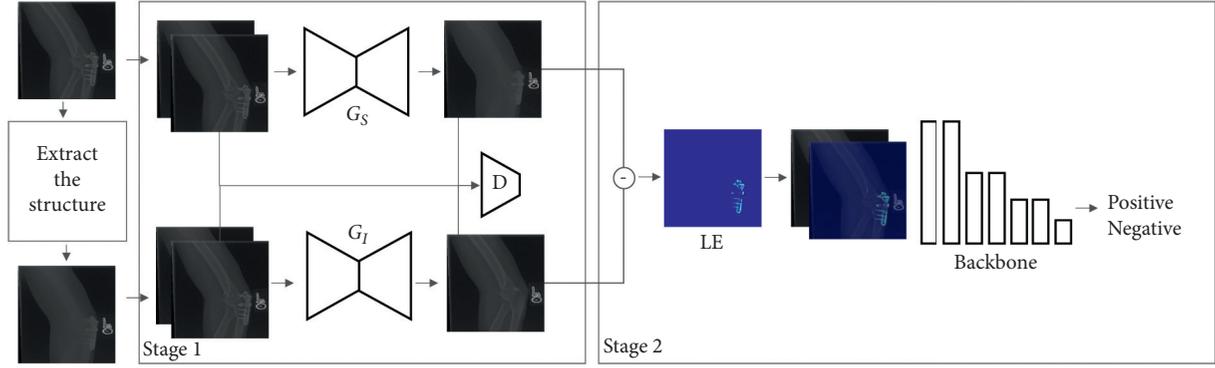


FIGURE 1: The overall structure of the two-stage ALENN module.

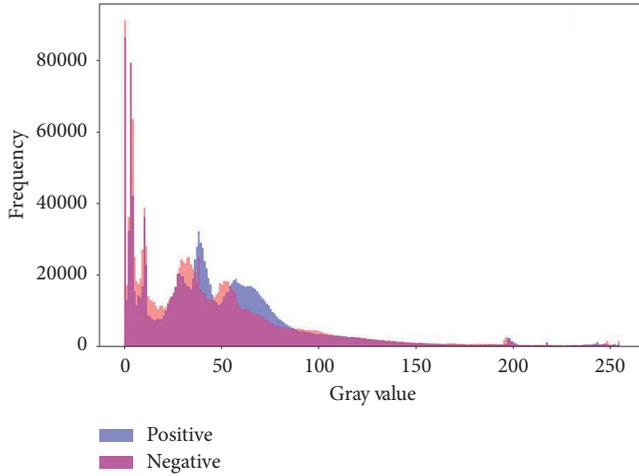


FIGURE 2: Data distribution in positive and negative samples.

3.1.2. GAN. In the architecture of this paper, the role of the generative adversarial network is to provide the lesion area in the elbow X-ray image for the final classification model. We additionally assume the MURA data: the distribution difference between the positive data and the negative data in the image results from the lesion. If the lesion area in the positive image is completely occluded, the part of the image that is not occluded at this time is similar in distribution to the negative data. The structure information obtained in Section 3.1.1 plays a significant role in GAN, which forces the generator to ignore the local interference caused by texture information, and then reconstruct the physiological structure information inside the image. And it fills and repairs the detailed texture on the basis of reconstructed structure information. More specifically, the image structure is used as the supervision information firstly, and the GAN is trained through the negative image repair task occluded by the random mask, so that the generator first fits the structure distribution of the bone and muscle in the negative data. On this basis, the original image is used as supervisory information to train GAN, so that the generator gradually fills in the detailed texture information on the basis of the structure. In the process of training the generator, the discriminator is also trained at the same time to judge the authenticity of the generated image and fight against the generator. Finally, the

positive data is used as the test data, and the lesion features in the image are repaired on the premise of preserving normal tissues, and then restored to pseudo-negative data.

This section follows the definitions of I and S in Section 3.1.1 and adds other definitions: generator G , discriminator D , and binarized matrix mask M . In addition, the image I is further expressed as $I = I_p \cup I_n$, where I_p represents the positive sample of the lesion in the image, and I_n represents a negative sample. Similarly, $S = S_p \cup S_n$. Based on this, the output of the GAN used to repair the image structure information in the first step can be expressed as

$$\hat{S} = G_{I \rightarrow S}(E(I_n, M), E(S_n, M)), \quad (5)$$

where $E(x, M) = x \odot M$ represents the occlusion operation; that is, the area with the value of 0 in M is occluded in the corresponding input data x . \odot denotes element-wise product. At the same time, the discriminator needs to judge the authenticity of the image and conduct adversarial training with the generator, and its output can be expressed as $T_S = D_S(S)$ and $\hat{T}_S = D_S(\hat{S})$.

In addition to training the generator to make the overall model capable to reconstruct structural information, it is also necessary to train the generator, so that the overall model has the ability to gradually repair texture information through structural information. And because this paper assumes “image = structure + texture,” after the texture information is supplemented and perfected on the basis of the structure information, it can be regarded as the overall restoration of the image. Therefore, the output of the generator $G_{S \rightarrow I}$ can be expressed as

$$\hat{I} = G_{s \rightarrow f}(E(I_n, M), \hat{S}). \quad (6)$$

Same as the repair structure stage, the discriminator still needs to predict the authenticity of the generator output image, and the output at this time is expressed as $T_I = D_I(I)$ and $\hat{T}_I = D_I(\hat{I})$.

3.1.3. Loss Function. The GAN in this paper requires two-step training. In other words, the training model firstly repairs the structure image with complex texture semantics removed. After the repair process is smooth, the training model will repair the detailed texture information based on

the structure. In the first training step, in order to fit the true distribution of the image structure S_n , we apply the idea of confrontation to the framework of the generative model. The

adversarial loss of the generator and the discriminator can be written as

$$L_{G,I \rightarrow \hat{S}} = \|S - \hat{S}\|_1, \quad (7)$$

$$\begin{aligned} L_{D,S} &= -\sum p(S) \log q(\hat{S}) = \mathbb{E}[\log(T_S)] + \mathbb{E}[\log(1 - \hat{T}_S)] \\ &= \mathbb{E}[\log(D_S(S))] + \mathbb{E}\left[\log\left(1 - D_S\left(G_{I \rightarrow \hat{S}}(E(I_n, M), E(S_n, M))\right)\right)\right]. \end{aligned} \quad (8)$$

Similar to the first training process, the loss function for the repair process of training GAN on texture details is as follows:

$$L_{G,\hat{S} \rightarrow \hat{I}} = \|I - \hat{I}\|_1, \quad (9)$$

$$\begin{aligned} L_{D,I} &= -\sum p(I) \log q(\hat{I}) = \mathbb{E}[\log(T_I)] + \mathbb{E}[\log(1 - \hat{T}_I)] \\ &= \mathbb{E}[\log(D_I(S))] + \mathbb{E}\left[\log\left(1 - D_I\left(G_{\hat{S} \rightarrow \hat{I}}(E(I_n, M), \hat{S})\right)\right)\right]. \end{aligned} \quad (10)$$

The two stages of adversarial losses are combined through hyper-parameters α and β , respectively, and go through the final optimization process:

$$L_S = \alpha L_{G,I \rightarrow \hat{S}} + (1 - \alpha) L_{D,S}, \quad (11)$$

$$L_I = \beta L_{G,\hat{S} \rightarrow \hat{I}} + (1 - \beta) L_{D,I}. \quad (12)$$

We set $\alpha = 0.8$ and $\beta = 0.85$ in this paper.

3.2. Lesion Fusion Classification Module. This section will introduce the second stage of ALENN, which is the medical image-assisted diagnosis module based on lesion information enhancement. If there is an auxiliary diagnosis model based only on the original data modeling, this model can well complete the auxiliary diagnosis task. However, due to the inexplicability of deep learning itself, the model can easily fall into a local optimal value. In addition, there also exist certain limitations on the convergence speed of the model, which depends on the model's initialization parameters and optimization strategies. Hence, the relative error between the positive data and the negative data and the similarity between the positive data are used to artificially amplify this distribution difference at this stage. More specifically, after the restoration of the positive data in the first stage, the corresponding pseudo-negative data is obtained, while the real negative data can still be regarded as negative data after being repaired. Based on this, we can enhance the lesion area in the data and then use it as the

modeling data for the auxiliary diagnosis model, making the model sensitive to differences in distribution.

The lesion fusion assisted diagnosis proposed in this paper is an overall framework, and the optimization of the specific network structure is beyond the scope of this paper. In other words, the method proposed in this paper is applicable to most classification networks. Therefore, we select the most basic model VGG-19 as the backbone of this framework. Next, we will gradually introduce the overall process of lesion information fusion in the second stage. In order to facilitate the presentation, we define the subsequent variables. Input image I , where W and H represent the width and height of the image, respectively. Set the sliding window set $D = \cup_i \text{window}_i$ for each input sample; the window size of the sliding window is w and h , which is a hyper-parameter set according to a priori. Similarly, the step size of the sliding window d is the hyper-parameter. Generally, we set $d < w$ so that each pixel in the input sample can get as many repair results as possible to reduce the impact of generation errors. From this, we can conclude that the number of sliding windows for an input sample is approximately $N = ((W - w) \times (H - h)) / d^2$. The lesion location algorithm based on sliding window is shown in Algorithm 1.

Consequently, under the supervision of category labels, we have completed the extraction of lesion information in medical images based on the difference of conditional probability distribution $P(I|\text{positive})$ and $P(I|\text{negative})$. Naturally, the artificial enhancement of this information makes the final auxiliary diagnosis model more sensitive to the difference in distribution between positive and negative

data. From another perspective, the enhancement of this lesion information can be seen as a priori attention to the lesion area to a certain extent. And this “prior” is not supervised, let alone man-made, but calculated from the difference in conditional probability distribution on image-level annotations.

4. The Experimental Results

4.1. Dataset and Training Setting. The MURA dataset is the largest public dataset of musculoskeletal image currently, jointly released by the Department of Computer Science, Medicine and Radiology of Stanford University. It contains a total of X training sets and X verification sets. The dataset selected 40,561 multi-view X-rays images of 12,173 patients from Stanford Hospital from 2001 to 2012 as sample data and was marked as normal or abnormal by professional radiologists. The result showed that 62% of the images were normal data while 38% were abnormal ones. The dataset consists of research types of finger, elbow, hand, humerus, forearm shoulder, and wrist. This paper only uses elbow data for experimental verification. And in order to show the experimental results fairly, the unified optimizer Adam is employed in this paper and the learning rate is initialized to 0.0001 with the batch size set to 16.

4.2. Evaluation Metric. In order to better evaluate the method proposed in this paper, we use peak signal-to-noise ratio (PSNR) coefficient and structural similarity (SSIM) to evaluate the performance of the repair module. The larger the value of the two indicators, the more similar the two images. And the accuracy, sensitivity, specificity, recall, F1-score, and Kappa score are applied to evaluate the performance of the classification module.

PSNR coefficient and SSIM are the most widely used objective measurement methods for evaluating repaired images. The calculation formula is as follows:

$$\text{PSNR} = 10 \cdot \log_{10} \left(\frac{\text{MAX}_I^2}{\text{MSE}} \right), \quad (13)$$

$$\text{SSIM}(x, y) = \frac{(2\mu_x\mu_y + c_1)(2\sigma_{xy} + c_2)}{(\mu_x^2\mu_y^2 + c_1)(\sigma_x^2\sigma_y^2 + c_2)}. \quad (14)$$

In this formula, μ_x is the average of x and μ_y is the average of y . σ_x^2 is the variance of x , σ_y^2 is the variance of y , and σ_{xy} is the covariance of x and y . $c_1 = (k_1 L)^2$ and $c_2 = (k_2 L)^2$ are used to maintain stability. L is the dynamic range of pixel values. $k_1 = 0.01$ and $k_2 = 0.03$. The range of structural similarity is 0 to 1. When the two images are exactly the same, the value of SSIM is equal to 1.

In addition, if FP, FN, TP, and TN represent the false positive rate, false negative rate, true positive rate, and true negative rate, respectively, the calculation formulas for accuracy, recall, precision, F1-score, and Kappa score are as follows:

$$\text{accuracy} = \frac{TP + TN}{TP + TN + FP + FN}, \quad (15)$$

$$\text{recall} = \frac{TP}{TP + FN}, \quad (16)$$

$$\text{precision} = \frac{TP}{TP + FP}, \quad (17)$$

$$F1 = 2 \cdot \frac{\text{precision} + \text{recall}}{\text{precision} + \text{recall}}, \quad (18)$$

$$\text{Kappa} = \frac{p_o - p_e}{1 - p_e}. \quad (19)$$

In this formula, p_o is the sum of the number of samples correctly classified in each category divided by the total number of samples, which is the overall classification accuracy. Assuming that the number of real samples in each category is a_1, a_2, \dots, a_n , respectively, and the predicted number of samples in each category is b_1, b_2, \dots, b_n , respectively, and the total number of samples is n , then $p_e = (a_1 \cdot b_1 + a_2 \cdot b_2 + \dots + a_n \cdot b_n) / (n \cdot n)$.

4.3. The Results of Lesion Enhancement. In order to verify the effectiveness of the first stage of the two-stage lesion enhancement classification method proposed in this paper, firstly it is necessary to qualitatively analyze the repair effect of the generated adversarial network. This section will perform visual analysis on the negative data and positive data of elbow data in MURA. Since the GAN in this paper is modeled based on negative data, in theory, GAN will approximate the distribution of negative data. On the contrary, when positive data is used as the input of GAN, GAN will repair the “abnormal” area in the positive data according to the negative data. The results of repair and visualization of negative and positive data are shown in Figure 3, in which the Difference indicates the difference between the original images and the restored ones and LE represents that the difference area is enhanced and displayed in the form of a heat map. The last column combines LE with the original images for more intuitive observation to the effectiveness of this method.

After qualitative analysis, we will quantitatively analyze the differences between the two data types, as shown in Table 1. It can be clearly seen that the repair effects of GAN are different for the two distributed images. However, this difference in distribution is directly reflected on the image and enhanced in this paper so as to provide clear guidance for auxiliary diagnosis of the two types of data in the next stage.

4.4. The Results of Classification. In order to prove the effectiveness and excellent adaptability of the method proposed in this paper, this section applies this method to each classification network. VGG-19, ResNet-50, DenseNet-121, and Inception-v3 are used to conduct related experiments and analyze the results. The experimental results are shown in Table 2.

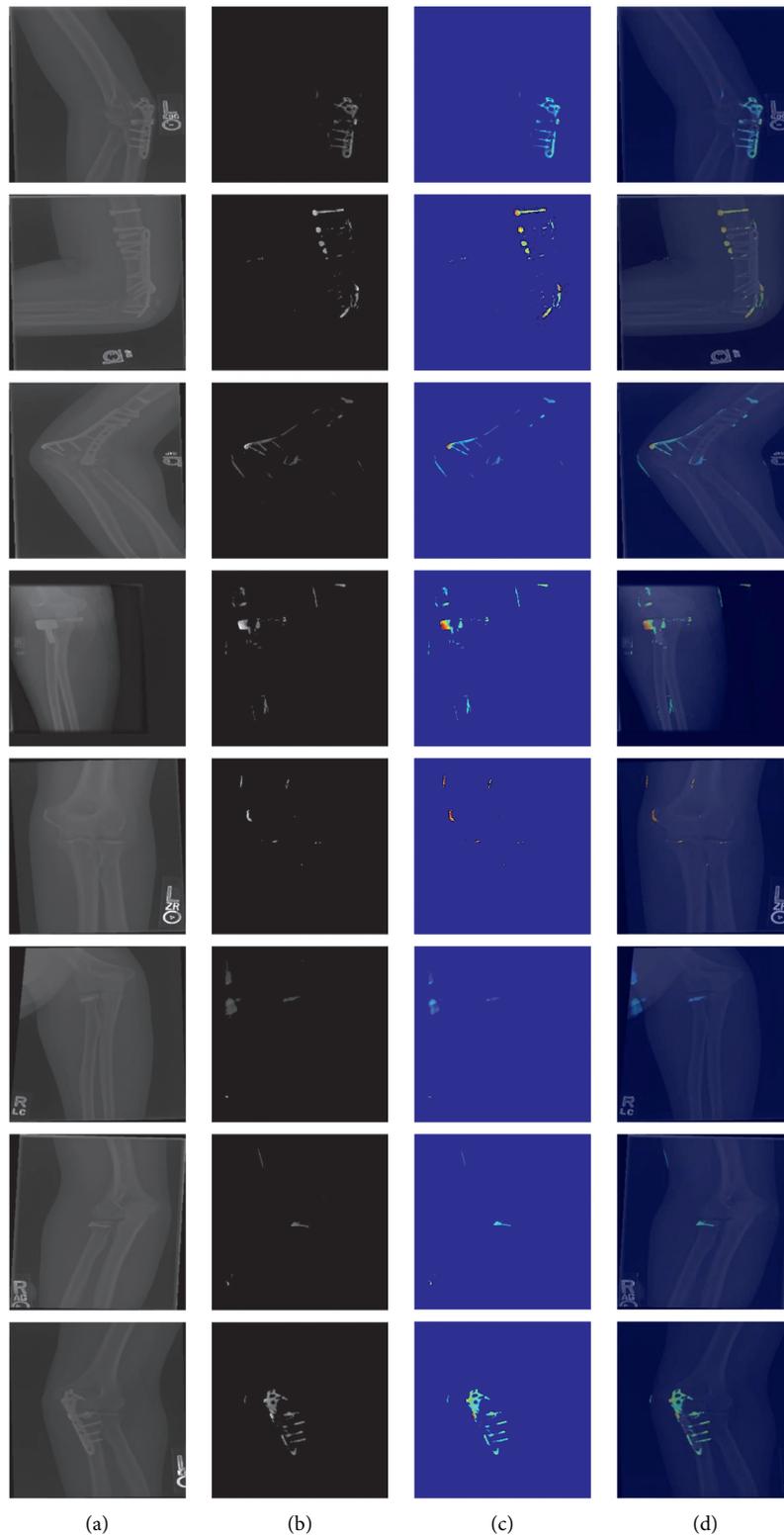


FIGURE 3: The results of repair and visualization of negative and positive data. (a) Original image. (b) Difference. (c) LE. (d) Color map.

It can be seen that the results are unsatisfactory compared to the method using only the classification network and the LE information. For each backbone, this method can increase the accuracy by about 1.67% on average. Moreover,

the method in this paper is the best one to lift VGG-19. The ROC curve of ResNet and VGG can explain this problem from another aspect, as shown in Figure 4. It is probably caused by the network structure of VGG itself. In other

TABLE 1: The differences between the two data types.

	PSNR	SSIM
Negative	34.7566	0.9708
Positive	34.6099	0.9657

TABLE 2: The experimental results.

Backbone	With LE	Accuracy	Precision	Recall	F1-score	Kappa
VGG-19	✓	82.8	76.82	94.47	84.73	65.5
		85.59	81.11	93.19	86.73	71.13
ResNet-50	✓	81.72	77.78	89.36	83.17	63.38
		83.23	78.97	91.06	84.59	66.39
DenseNet-121	✓	81.51	78.54	87.23	82.66	62.96
		83.01	82.77	83.83	83.3	66.01
Inception-v3	✓	82.15	78.36	89.36	83.5	64.24
		83.01	78.68	91.06	84.42	65.96

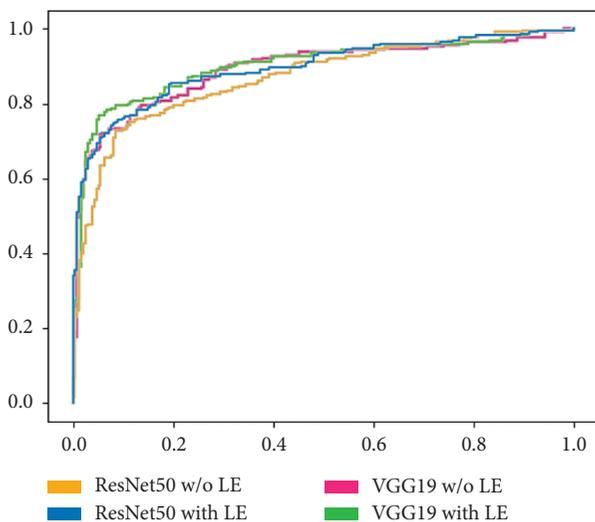


FIGURE 4: The ROC curve of ResNet and VGG.

words, compared to VGG with a simple straight-line connection, ResNet and other networks have a more complicated network structure, such as short connections. It may help the modeling process to better find the lesion area to a certain extent. However, more in-depth theoretical proofs and related experiments are required to prove this.

5. Conclusion

With the gradual improvement of deep neural network technology recently, there are more methods focusing their researches on exploring the hidden information inside data. Taking computer vision as an example, many studies have bundled the task of exploring the semantic information inside the image with the optimization of the neural network structure. Among them, medical imaging has become one of the important objects of image semantic mining due to its high semantic consistency. However, the interpretability of deep learning itself has always been a topic still under studying; that is, the characteristics learned by neural

networks cannot be explained intuitively. In other words, the correlation between the optimization of neural network structure and the effectiveness of semantic information mining remains to be verified. Therefore, this paper proposes an Adversarial Lesion Enhancement Neural Network for Medical Image Classification, which takes the extraction of this hidden semantics as a separate stage, separating it from the auxiliary diagnosis model. The purpose of this is to (1) clearly show the effectiveness of semantic information extraction, (2) use this method as a portable auxiliary diagnostic module with high adaptability, and (3) complete the lesion positioning with only coarse-grained labels. Finally, this paper proves the effectiveness of the first stage structure-based lesion adversarial inpainting module on the public dataset MURA. Meanwhile, on this basis, it is proved that the combined use of two-stage modules can improve the auxiliary diagnosis model. However, there is also a shortcoming in this method, that is, the relatively high time complexity of the sliding window, which is also one of the optimization directions of this research in the future.

Data Availability

The X-ray images used to support the findings of this study have been deposited in the MURA repository.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

References

- [1] I. Stuart, E. H. Y. Weinstein, and I. Sylvia, "Musculoskeletal research," 2014, <http://www.boneandjointburden.org/2014-report>.
- [2] D. Banga and W. Peter, "Abnormality detection in musculoskeletal radiographs with convolutional neural networks(ensembles) and performance optimization," *CoRR*, vol. 217, 2019.
- [3] S. Macias-Velasquez and A. Yolanda, "Working hours, burnout and musculoskeletal discomfort in middle and senior management of mexican industrial sector," *IEEE Access*, vol. 8, pp. 48607–48619, 2020.
- [4] K. Wang, L. Wang, Z. Deng, C. Jiang, W. Niu, and M. Zhang, "Influence of passive elements on prediction of intradiscal pressure and muscle activation in lumbar musculoskeletal models," *The Computer Methods and Programs in Biomedicine-Update*, vol. 177, pp. 39–46, 2019.
- [5] T. Tien-Loc Le, Lo-Yi Lin, C.-M. Lin, and F. Chao, "A k-means interval type-2 fuzzy neural network for medical diagnosis," *International Journal of Fuzzy Systems*, vol. 21, no. 7, pp. 2258–2269, 2019.
- [6] S. Raouafi, S. Achiche, M. Begon, A. Sarcher, and M. Raison, "Classification of upper limb disability levels of children with spastic unilateral cerebral palsy using k-means algorithm," *Medical & Biological Engineering & Computing*, vol. 56, no. 1, pp. 49–59, 2018.
- [7] M. Al-Maitah, "Analyzing genetic diseases using multimedia processing techniques associative decision tree-based learning and hopfield dynamic neural networks from medical images," *Neural Computing and Applications*, vol. 32, no. 3, pp. 791–803, 2020.

- [8] H. Faris, M. Habib, M. Faris, M. Alomari, and A. Alomari, "Medical speciality classification system based on binary particle swarms and ensemble of one vs. rest support vector machines," *Journal of Biomedical Informatics*, vol. 109, Article ID 103525, 2020.
- [9] C. Guo, J. Zhang, Y. Liu, Y. Xie, Z. Han, and J. Yu, "Recursion enhanced random forest with an improved linear model (RERF-ILM) for heart disease detection on the internet of medical things platform," *IEEE Access*, vol. 8, pp. 59247–59256, 2020.
- [10] M. Gadermayr, B. Mara Klinkhammer, B. Peter, and D. Merhof, "CNN cascades for segmenting sparse objects in gigapixel whole slide images," *Computerized Medical Imaging and Graphics*, vol. 71, pp. 40–48, 2019.
- [11] X. Cao, M. Liu, F. Zhai et al., "Comparison of different registration methods and landmarks for image-guided radiation therapy of pulmonary tumors," *BMC Medical Imaging*, vol. 19, no. 1, pp. 1–46, 2019.
- [12] P. Rajpurkar, J. Irvin, A. Bagul et al., "MURA: large dataset for abnormality detection in musculoskeletal radiographs," in *Proceedings of the 1st Conference on Medical Imaging with Deep Learning (MIDL 2018)*, Amsterdam, The Netherlands, 2018.
- [13] A. Amirkhani, M. R. Mosavi, K. Mohammadi, and E. I. Papageorgiou, "A novel hybrid method based on fuzzy cognitive maps and fuzzy clustering algorithms for grading celiac disease," *Neural Computing and Applications*, vol. 30, no. 5, pp. 1573–1588, 2018.
- [14] J. S. B. Sang Gil Lee, H. Kim, H. Jung, and S. Yoon, "Liver lesion detection from weakly-labeled multi-phase ct volumes with a grouped single shot multibox detector," in *Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention*, Granada, Spain, 2018.
- [15] N. Dong, Li Wang, L. Xiang, S. Zhou, E. Adeli, and D. Shen, "Difficulty-aware attention network with confidence learning for medical image segmentation," in *Proceedings of the The Thirty-Third AAAI Conference on Artificial Intelligence, AAAI 2019*, pp. 1085–1092, AAAI Press, Honolulu, Hawaii, USA, 2019.
- [16] X. Wang, Y. Peng, Le Lu, Z. Lu, M. Ronald, and Summers, "Tienet: text-image embedding network for common thorax disease classification and reporting in chest x-rays," in *Proceedings of the 2018 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2018*, pp. 9049–9058, IEEE Computer Society, Salt Lake City, UT, USA, 2018.
- [17] Qi Wang, J. Qiu, Y. Zhou, R. Tong, D. Gao, and Ju Gao, "Automatic severity classification of coronary artery disease via recurrent capsule network," in *Proceedings of the IEEE International Conference on Bioinformatics and Biomedicine, BIBM 2018*, pp. 1587–1594, IEEE Computer Society, Madrid, Spain, 2018.
- [18] P. Kumar Mall, P. K. Singh, and D. Yadav, "GlcM based feature extraction and medical x-ray image classification using machine learning techniques," in *Proceedings of the 2019 IEEE Conference on Information and Communication Technology (CICT)*, Allahabad, India, 2019.
- [19] P. Rajpurkar, J. Irvin, K. Zhu et al., "Chexnet: radiologist-level pneumonia detection on chest x-rays with deep learning," *CoRR*, vol. 225, 2017.
- [20] A. F. M. Saif, C. Shahnaz, W.-P. Zhu, and M. Omair Ahmad, "Abnormality detection in musculoskeletal radiographs using capsule network," *IEEE Access*, vol. 7, pp. 81494–81503, 2019.
- [21] L. Oakden-Rayner, J. Dunnmon, G. Carneiro, and R. é Christopher, "Hidden stratification causes clinically meaningful failures in machine learning for medical imaging," in *Proceedings of the ACM CHIL '20: ACM Conference on Health, Inference, and Learning*, pp. 151–159, Toronto, Canada, 2020.
- [22] Y. Ren, X. Yu, R. Zhang, T. H. Li, S. Liu, and Li Ge, "Structureflow: image inpainting via structure-aware appearance flow," in *Proceedings of the 2019 IEEE/CVF International Conference on Computer Vision, ICCV 2019*, pp. 181–190, IEEE, Seoul, Korea, 2019.
- [23] Li Xu, Q. Yan, Y. Xia, and J. Jia, "Structure extraction from texture via relative total variation," *ACM Transactions on Graphics*, vol. 31, no. 6, p. 139, 2012.